EQUIVALENCE CLASS FORMATION IN LANGUAGE-ABLE AND LANGUAGE-DISABLED CHILDREN

JEANNE M. DEVANY, STEVEN C. HAYES, AND ROSEMERY O. NELSON

AUBURN UNIVERSITY, UNIVERSITY OF NEVADA-RENO, AND
UNIVERSITY OF NORTH CAROLINA-GREENSBORO

Stimulus equivalence seems to have relevance to the study of semantics and of language more generally. If so, there may be a relation between language use and the demonstration of stimulus equivalence. This was examined in three groups of children ranging in chronological age and matched on a conventional measure of mental age: normally developing preschoolers, retarded children who used speech or signs spontaneously and appropriately, and retarded children who did not. All children were taught a series of four related discriminations and were then tested to determine if classes of equivalent stimuli had formed. All of the language-able children (retarded and normal) formed equivalence classes, whereas none of the language-disabled children did so. Although the exact nature of the relation between stimulus equivalence and language remains to be clarified, these results support the view that stimulus equivalence is a phenomenon with relevance to language.

Key words: stimulus equivalence, conditional discrimination, language, verbal behavior, children

---

In traditional views of language, much is made of the symbolic nature of words, but relatively little work has been done to show why or how words come to function as symbols. Instead, the literature has asserted that words do act as symbols and has traced their use. Verbal humans are said to be able to "manipulate symbols" (Clark & Clark, 1977), to "map words onto internal concepts" (Nelson, 1974), or to use words to "refer" to objects, events, or relations (Premack, 1976). Exactly what constitutes a symbol and what gives rise to symbolic relations in verbal humans is rarely addressed. For instance, the textbook quoted above by Clark and Clark repeatedly refers to the symbolic nature of language, but fails even to include the word "symbol" in its index. It is as if the origin or nature of symbolic activity per se need not be explained.

Behavioral perspectives should be well positioned philosophically to address the ontogeny of symbolic relations. Previous views of the nature of meaning, however, have restricted the scope of behavior-analytic investigation. From a Skinnerian viewpoint, the "meaning" of a word to a listener is established through direct contingencies embedded in interactions with the verbal community (Skinner, 1957). The symbolic meaning of the arbitrary vocal stimulus "bread," for example, is found through observation of the behavior it occasions, based on the direct contingencies of reinforcement in which it participates. Other than the fact that these contingencies are largely social, the meaning of a word is thought to be established in a manner functionally identical to stimulus-control processes readily seen in infrahumans. From the standpoint of the listener, this analysis essentially views symbols as discriminative stimuli and not of special interest in their own right.

Symbols used by humans, however, appear to be much more flexible than discriminative stimuli as typically conceived. For example, in a Quebecois human child, the English written word "bread" or the French written word "pain," their spoken counterparts, pictures of bread, and loaves of bread all enter into a rich network of relations in which each may (in a sense) stand for the others. This seems to be a defining property of symbols and is consistent with the word's etymology. The word "symbol" comes from root words meaning "together" and "to throw." Symbols, then, are

stimuli that are "thrown together" with other stimuli. The relation between a symbol and a referent seems necessarily bi-directional. A word "stands for" another event only if the event "is called" the word.

There is one behavioral process, stimulus equivalence, that appears to relate quite closely to the issue of symbolic activity (Sidman & Tailby, 1982). When humans are taught a series of related conditional discriminations, the component stimuli of the discriminations often become related to each other in new ways that had not been explicitly taught in training (e.g., Sidman, Cresson, & Willson-Morris, 1974). For example, if a child is taught to match A to B and then A to C, it is likely that the child will, without additional training, be able to match B to A and C to A and, perhaps most importantly, match B to C.

We speak of an equivalence class if the stimuli in the class show the three defining relations of reflexivity, symmetry, and transitivity (Sidman & Tailby, 1982). In matching-to-sample procedures, reflexivity is generalized identity matching—matching a novel stimulus to itself under conditions of no reinforcement. That is, if an organism is presented with bread, it will select bread from an array of items even in the absence of programmed reinforcement of that choice. Symmetry refers to the functional reversibility of the conditional relation: if A, then B; and if B, then A. In the presence of the printed word BREAD, a loaf of bread is selected; and, in the presence of a loaf of bread, the printed word BREAD is selected. This reversibility must be demonstrated in the absence of direct reinforcement to be considered symmetry (Sidman, Rauzin, Lazar, Cunningham, Tailby, & Carrigan, 1982). To demonstrate transitivity, at least three stimuli are required. If after the relations "if A, then B" and "if B, then C" have been taught, the relation "if A then C" emerges without additional training, transitivity has been demonstrated. For example, if the child has been taught to select bread upon hearing "bread" and to select the printed word BREAD upon seeing bread, when the child selects BREAD upon hearing "bread," without any additional teaching, then transitivity has been demonstrated.

In the context of stimulus equivalence, a "symbol" and its "referents" form a class of functionally substitutable elements. The re-

lation between a symbol and its referent is not a unidirectional conditional relation (although the members of the class are conditionally related to each other); the relation is functionally reversible. The relations among the members of an equivalence class appear to approximate what psycholinguists and others mean when they say that a word represents or "stands for" its referent in a way that a conditionally related response does not.

If equivalence classes have to do with verbal or symbolic activity, and if symbolic stimuli differ from more typical forms of discriminative stimuli, we would expect that it would be easy to demonstrate the formation of equivalence classes with humans and difficult to do so with nonhumans. This is exactly what has been found in the research done to date. The formation of classes of equivalent stimuli has been demonstrated using a wide variety of human subjects and materials (Dixon, 1976; Dixon & Spradlin, 1976; Gast, VanBiervliet, & Spradlin, 1979; Mackay & Sidman, 1984; Sidman, 1971; Sidman et al., 1974; Sidman & Tailby, 1982; Spradlin, Cotter, & Baxley, 1973; Spradlin & Dixon, 1976; VanBiervliet, 1977; Wetherby, Karlan, & Spradlin, 1983). Although conditional relations have been demonstrated in a variety of infrahumans, including dolphins (e.g., Herman & Thompson, 1982), rats (e.g., Lashley, 1938), pigeons (e.g., Edwards, Jagielo, & Zentall, 1983), and monkeys (e.g., Nissen, 1951), these relations have not yet been shown to result in stimulus equivalence as they do in most humans. To date, there has been no success in unequivocally demonstrating the formation of an equivalence class in any infrahuman, including the higher primates, although efforts have been made (Sidman et al., 1982). Further, researchers have had considerable difficulty demonstrating transitive transfer in chimpanzees and other infrahumans even under conditions of direct reinforcement (Fouts, Chown, & Goodin, 1976; Kendall, 1983).

A type of "transitive" responding has recently been demonstrated in monkeys (D'Amato, Salmon, Loukas, & Tomie, 1985), but it did not require symmetrical responding. This clearly distinguishes it from stimulus equivalence as currently defined, and may make their results more amenable to direct conditioning explanations. Similarly, very complex classes of stimuli can be established

in chimpanzees (Savage-Rumbaugh, 1984), but no direct tests of symmetry or transitivity have been done to see if these classes function as equivalence classes (Savage-Rumbaugh, personal communication, May 1985). These various findings suggest that the ability to form equivalence classes is related to language development, not because language and stimulus equivalence both do not occur in infrahumans—with proper training they may—but because both are known to occur with relative ease in most humans, and are at least difficult to show in infrahumans.

Thus, the relations seen among stimuli in an equivalence class seem to parallel the symbolic relations commonly said to be characteristic of language. If so, the ability to form equivalence classes should be related to language acquisition or use; however, no experiments relating language use and equivalence classes have yet been published. One approach to the problem is to compare the performances of normally developing children with the performances of language-impaired children on an equivalence test. If the normally developing children were able to form equivalence classes and the language-deficient children were not, the data would provide further support for the view that the ability to form classes of equivalent stimuli is related to language. Additional research could begin to address the precise nature of that relation.

The present project addressed the question: Is there a difference between normal and language-deficient children in performance on tests of equivalence? Three groups of children were used: normally developing children, retarded children demonstrating some expressive speech, and retarded language-deficient children. The three groups were matched with respect to conventional measures of mental age. Each of the children was taught a series of conditional discriminations and was subsequently tested to determine if an equivalence class had formed.

METHOD

Subjects and Subject Identification

Twelve children, four in each group, whose measures of mental age ranged from 14 to 36 months, served as subjects. The first group consisted of normally developing preschoolers

recruited from the UNC–G Child Care Center. The second and third groups were composed of retarded children. All of these children were enrolled in educational programs at the Henry Wiseman Kendall Center in Greensboro.

All of the normally developing preschoolers had speech skills that were generally consistent with their chronological ages. No formal assessment of their speech and language skills was done; however, in the training and testing sessions no abnormalities of speech or language were noted. In addition, no abnormalities were noted by the classroom teacher, nor were any observed during in-class observation by the experimenter.

Half of the retarded children engaged in speech outside of language-training sessions. All of these children spoke in complete, albeit brief, sentences when prompted and often spontaneously asked for desired items or commented on events in the classroom. Two of these children had articulation problems, however (Carl and Allen, names changed to protect confidentiality). The other retarded children lacked functional speech or language skills. None of these children spontaneously used words, signs, or picture boards. Two of the children were echolalic, repeating words or phrases without evidence of comprehension (Debbie and Andrew), and two of them uttered vowel sounds (Craig and Barb).

The retarded children were classed into language and no-language groups on the basis of converging categorization from three independent observers. The speech pathologist at the retarded children's institution was asked to categorize all of the children in two preschool classes into one of two categories: possessing functional speech or sign skills (used in communication, even if poorly articulated), or as lacking functional speech or sign skills. This was done before the study began. She was not told the purpose of the study other than that it was an investigation of the ease of concept learning in children of varying levels of language and cognitive skills. The experimenter (JMD) observed each of the subjects in the classroom for a minimum of one and a half hours prior to the onset of the project. Children were categorized on the basis of whether they used signs or speech during this period. In addition, the reliability observer used throughout the experiment was an ad-

Table 1

Subject characteristics.

	Mental age (in months)	Chronological age (in years-months)
Normal children		
Alex	30	2-6
Bobby	37	2-11
Claire	20	2-6
Diane	19	2-1
Retarded/language children		
Allen	31	3-7
Beth	36	4-4
Carl	20	3-3
David	19	2-8
Retarded/no-language children		
Andrew	30	4-1
Barb	36	4-4
Craig	18	4-4
Debbie	14	2-7

Fig. 1. Examples of the stimulus figures used in this experiment.

vanced graduate student with many years of experience working with retarded and language-impaired children. He observed each of the subjects at least once for approximately an hour and was not informed of their previous classification. After the session was completed, he was asked to categorize the child on the basis of his or her behavior in the session. In all cases, the three observers were in agreement.

The children in the three groups were matched for mental age on the basis of an individual intelligence test (the Stanford-Binet Intelligence Scale, Form L-M, or the Bayley Scales of Infant Development). The retarded children were all assessed by personnel at their facility. The normal preschoolers were assessed by the experimenter. There were four levels of mental age for each group, ranging from 14 to 36 months.

The children's names were changed to protect confidentiality. Each child was given a name beginning with the same letter as the names of his or her mental-age-matched controls. The mental-age scores and chronological ages of each child are presented in Table 1.

Stimuli

In the training phase, the children were taught four conditional discriminations: if A, then B; if D, then E; if A, then C; if D, then F. Four tasks were learned by each child (A-
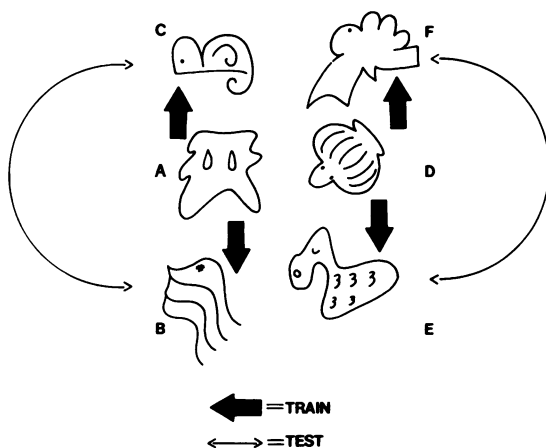
B, D-E, A-C, and D-F). The tasks consisted of matching made-up animal-like figures using a matching-to-sample format. On each trial, the A or D stimulus was presented as a sample with either B and E or C and F as comparisons. Each child was trained and tested using a different stimulus set, made by randomly selecting from a pool of items. All of the stimulus figures were colored with watercolor magic markers. Six colors were used: red, brown, green, purple, yellow, and orange. Each item was colored a different color. Color assignment was random, except that all six colors had to be used in each set of stimuli. A black-and-white example of the stimuli is shown in Figure 1.

In a given trial, the three stimuli (one sample and two comparisons) were presented on 8½ × 11-in. sheets of white paper. When the long side of the sheet was placed horizontally, the sample stimulus was at the top center of the page and the two comparisons were in the bottom half of the page, each 1.5 in. from the edge of the sheet. The left–right order of presentation of the comparison stimuli was counterbalanced across trials to prevent the child from responding correctly on the basis of stimulus position. Each sheet contained the stimuli for one trial.

The materials used in the test phase were identical to those used in the training phase, except that the sample stimuli were stimuli that previously had been comparisons during the conditional discrimination training. Equivalence would be indicated by matching

between B and C, inasmuch as A had been the matching sample for both, and between E and F, both of which had been matched to D. Ten trials of the test phase were devoted to each of four types of problems: either B or E as a sample with C and F as comparisons; either C or F as a sample with E and B as comparisons.

Setting

The sessions for the retarded children were held at a small table in an unused classroom. The child always sat at the table, which was placed against a wall. The experimental materials were placed on the table in front of the child or were held up off the table slightly to permit easier viewing. Sessions for the normal children were held in a small office. The child and the experimenter sat on a rug on the office floor. The stimulus sheets were presented by placing them on the floor directly in front of the child or by holding them slightly off the floor.

Procedure

Each child was taught individually, and the same training sequence was used for all. First, the child was taught to select B in the presence of A (A-B). Then the child was taught to select E in the presence of D (D-E). Then these two tasks were mixed; the stimulus cards from both sets were mixed together and presented in a random order. Once this task was mastered, training on A-C was begun. Then D-F was taught. Once D-F was mastered, A-C and D-F were mixed and presented to the child. When the child reached the mastery criterion on this mixed task, the stimulus items from all four tasks were mixed and presented. Once the child reached the mastery criterion on this final task, the test trials were presented. The mastery criterion used throughout training was 9 of 10 consecutive unprompted responses correct.

At the beginning of each session, the experimenter greeted the child and spent several moments conversing. Even if the child was nonverbal, the first portion of the session was spent talking to the child in order to set a relaxed and pleasurable tone. The experimenter started off the experimental procedure by saying, "I have some things I would like you to help me with. Let's see if you can help me. I also have some things to play with and

we will play with those as well." For the retarded/no-language children this was shortened to "Let's do some work," which was the standard cue used for these children. The first task (A-B) was then presented. At the beginning of each trial, the experimenter pointed to the sample and said, "Touch the one that goes with this one."

For the normal children, correct responses during the training phase produced one of several consequences including praise, blowing soap bubbles, and singing. Correct responses by the retarded children produced brief access to tiny flashlights, soap bubbles, balloons, juice, and cheese crackers. When necessary, physical prompting (guiding the child's hand to the correct choice) and visual prompting (placing the experimenter's finger on the correct choice) were used as teaching aids. Visual prompting was used in a trial if the child had failed to make a response during the previous trial or if the previous two responses had been incorrect. If the child's response following visual prompting was correct, the visual prompt was faded by placing a finger 2 in. away from the correct stimulus on the next trial, 6 in. away on the next trial, and then removing the prompt entirely. Physical prompting was used if the visual prompt failed to produce correct responding. After guiding the child's hand to the correct choice, the physical prompt was faded on subsequent trials. First, very light pressure was used to guide the child's hand, then the experimenter's hand shadowed the child's without any guidance, then the child's hand was touched briefly after the choice, and finally the prompt was removed entirely. If an incorrect response occurred, the previous level of assistance was reinstituted.

Initially, all correct responses led to the delivery of one of the consequences. At the end of the training phase, when the conditional discriminations were mixed and presented, the schedule was gradually thinned until a programmed consequence was delivered only after every three or four correct responses. This was done to equate the rate of reinforcement to that used in the testing phase.

In the test phase, 40 trials were presented. Rewards contingent on responses consistent with the formation of an equivalence class were not delivered during testing. The composition of these trials has already been de-

Table 2

Number of sessions to final mastery criterion.

Alex	2	Bobby	2
Allen	7	Beth	6
Andrew	16	Barb	8
Claire	16	Diane	21
Carl	7	David	11
Craig	14	Debbie	16

Table 3

Interobserver agreement during multiple sessions when reliability measures were obtained.

Number of trials	Agreement
67	88
60	100
150	96
260	100
200	100
120	100

scribed. The four trial types were randomly intermixed. After every third or fourth response (correct or incorrect), the child was praised or one of the other programmed consequences was delivered for cooperation, good sitting, and the like. If the child asked for explicit feedback about a response, the experimenter said, "In this part of the game, I must be very quiet. I think you are doing a good job of working on this."

The data were collected over several sessions with each child. Most sessions lasted approximately 20 min, and no session ran longer than 30 min. The number of sessions required to reach the final training mastery criterion is presented in Table 2. The equivalence test was always conducted immediately after the child reached the final mastery criterion. For some children, this occurred near the end of a session; for others this occurred near the beginning.

Recording and Reliability

Behavior in each trial was scored as "correct," "incorrect," or "no response." The experimenter was the primary data collector. A correct response was defined as touching the correct comparison stimulus while refraining from touching the incorrect comparison or the sample stimulus. An incorrect response was defined as touching the incorrect comparison, touching the sample, touching both the correct comparison and the sample or incorrect comparison, or touching another part of the stimulus sheet. A no response was defined as any other behavior. In the testing phase, the "correct" response was the response that would be expected if an equivalence class had formed.

Reliability data were collected in 20% of the sessions, distributed across children. These data were collected by a trained graduate student who was familiar with the general nature of the project but was unfamiliar with the specific hypothesis. The rater sat in a po-

sition from which he could not observe the experimenter's data sheet. During sessions in which the observer was present, the experimenter paused briefly after each trial before delivering the consequence, permitting the observer to record the data without knowledge of the experimenter's scoring.

Reliability was calculated on a trial-by-trial basis using the formula [Agreements/(Agreements + Disagreements)] × 100. An agreement was scored if the two observers both recorded a response as correct or as incorrect. For the purposes of reliability computations, a prompted trial was considered correct. Interobserver agreement per session ranged from 88% to 100%. These data are summarized in Table 3.

RESULTS

Conditional Discrimination Training

The individual data grouped by measured mental age are presented in Figures 2, 3, 4, and 5. In each figure, the data for the matched normal child, retarded/language child, and retarded/no-language child are presented. The data are graphed as the percentages of unprompted correct responses in blocks of 10 consecutive trials.

An inspection of the individual data shows that performance varied among the three subject groups. The retarded/language and normal groups consistently required fewer trials to complete the conditional discrimination training than did children in the retarded/no-language condition. In the normal group, Alex required 95 trials to complete the training, and Bobby required 107 trials; Claire required 185 trials, whereas Diane required 273 trials. In the retarded/language group, Allen required 277 trials, Beth required 223 trials, Carl required 227 trials, and David required
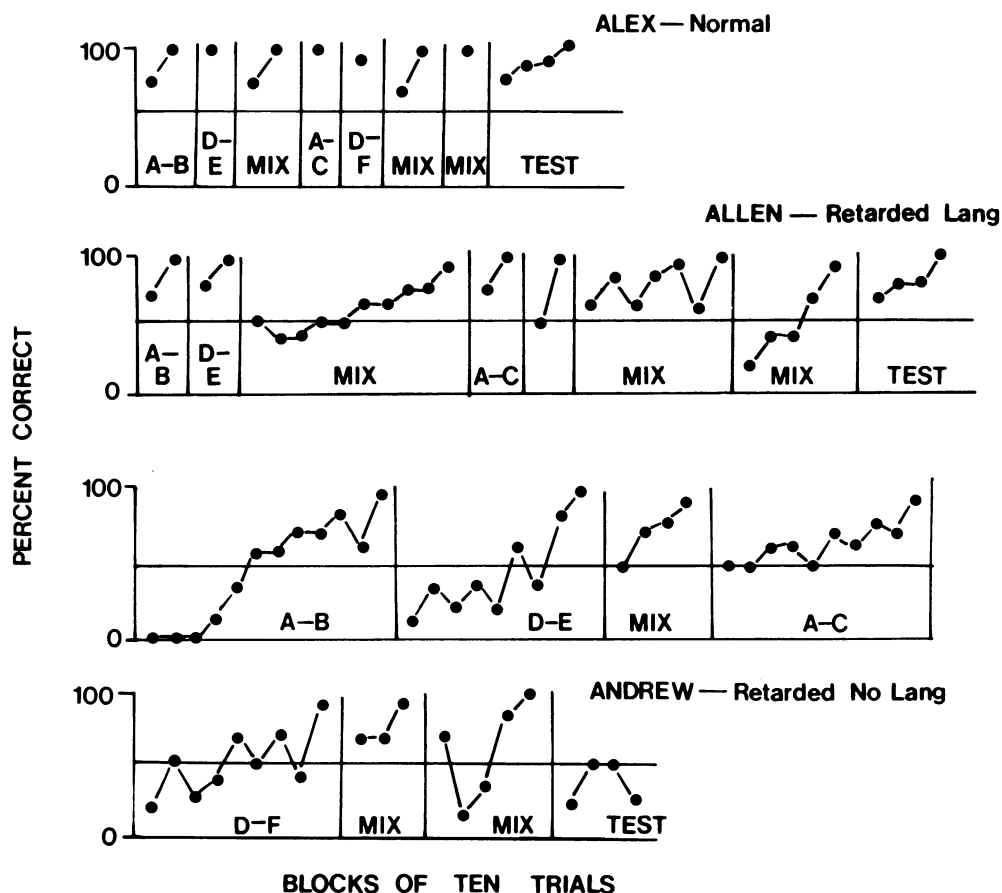
Fig. 2. Individual training and testing data for Alex, Allen, and Andrew (MA = approximately 30 months). The data are presented as the percentage of correct unprompted trials (vertical axis) across blocks of 10 trials (horizontal axis).

264 trials. In the retarded/no-language group, Andrew required 507 trials, Barb required 280 trials, Craig required 370 trials, and Debbie required 750 trials.
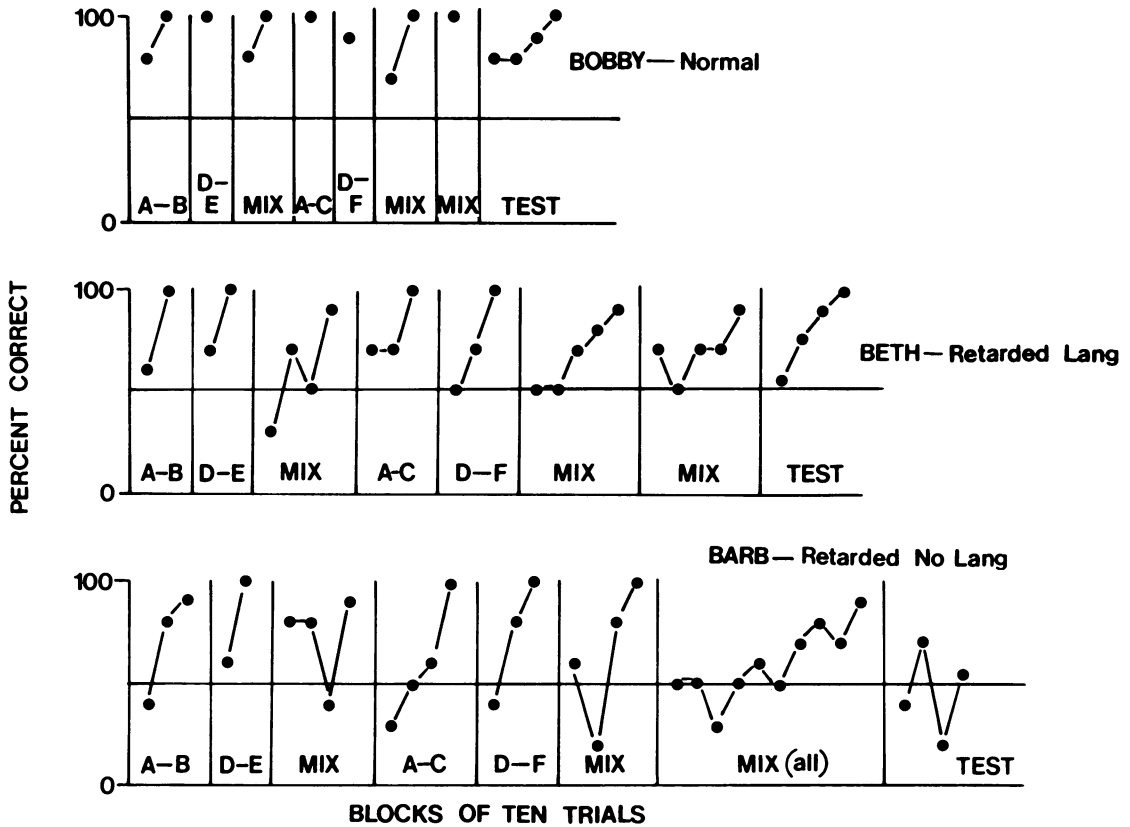
An analysis of variance indicated a significant difference among the three groups in the number of trials needed to complete the conditional discrimination training, $F(2, 9) = 6.34$, $p < .019$. A Newman-Keuls post-hoc analysis (Ferguson, 1976) revealed no significant difference between the normal and the retarded/language group, but did between these groups and the retarded/no-language group ($p < .05$).

There was also a difference among the three groups in the number of prompts used in the conditional-discrimination training. The mean number of prompts (visual and manual) used per child was 29 for the normal group, 40 for the retarded/language group, and 184 for the retarded/no-language group.

Equivalence Test

The test data are presented in different form in Figure 6. Each graph represents one subject. The columns in each individual graph represent the number of "no responses" made by the child during each block of 10 trials. As "no responses" do not allow us to determine if an equivalence class has formed, the percentages of correct responding were calculated as the number of correct responses divided by the total number of responses in that block. Each row of graphs represents the data from one group with respect to language. Each column of graphs represents the data for those children matched for mental-age scores (across groups).

Fig. 3. Individual training and testing data for Bobby, Beth, and Barb (MA = approximately 36 months). The data are presented as the percentage of correct unprompted trials (vertical axis) across blocks of 10 trials (horizontal axis).

The data confirmed the prediction that the language-able children would perform better than the language-deficient children on the stimulus-equivalence test. Calculated as just described, the average correct responding in the test phase in the normal group was 84.5%. In the retarded/language group, the average percentage of correct responding was 78.25%. In the retarded/no-language group, the average correct responding was 44.5%—very close to chance level (50%). No increase across trials in the number of no responses was seen in any group. The effects were quite consistent across children with widely varying measured mental age. Every one of the subjects in the normal and retarded/language groups showed high percentages of correct responding, whereas the performances of all of the children in the retarded/no-language group remained near chance level throughout. The exception to this was Craig (column C, third

graph). His performance deteriorated during the test until correct responding was at 0%. Notations made on the data sheets during the test phase indicated that he consistently touched the center (white space) of the stimulus sheets rather than consistently choosing the incorrect comparison stimulus.

All of the normal and retarded/language children improved from the first to the last half of the testing period. For the normal children, the mean percentage of correct responding in the first half was 77.75%, and the mean for the second half was 95.5%. For the retarded/language children, the mean percentage of correct responding in the first half was 69.75%, and during the second half, 88%. For the retarded/no-language group, mean correct responding during the first half was 46.25%, and during the second half, 39.25%. Improvement in performance during testing without reinforcement procedures has previ-
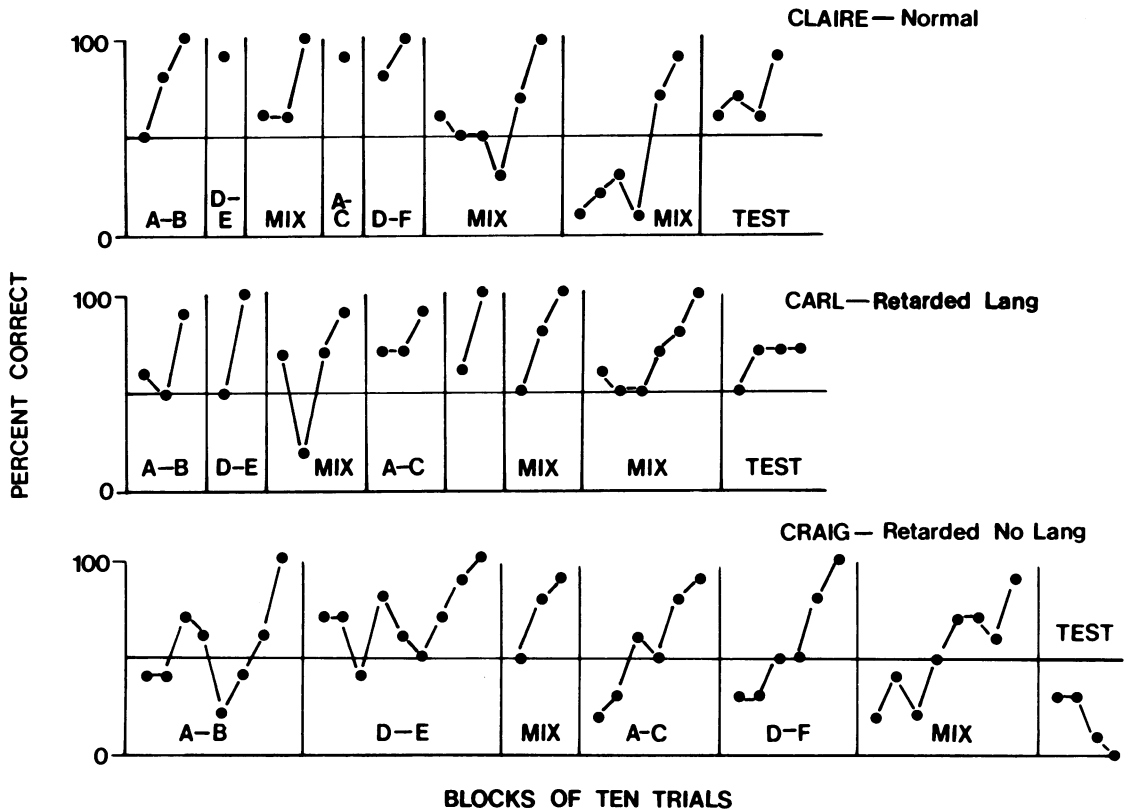
Fig. 4. Individual training and testing data for Claire, Carl, and Craig (MA = approximately 20 months). The data are presented as the percentage of correct unprompted trials (vertical axis) across blocks of 10 trials (horizontal axis).

ously been found in the equivalence literature (e.g., Sidman, Kirk, & Willson-Morris, 1985).

A one-way ANOVA on the percentage of correct responding confirmed that the language-able children performed significantly better than the language-deficient children on the stimulus equivalence test, $F(2, 9) = 18.51$, $p < .0006$. A Newman-Keuls analysis revealed no significant difference between the normal and the retarded/language group, but did between the retarded/no-language and the normal ($p < .01$) and the retarded/language ($p < .01$) groups.

DISCUSSION

During the test, each language-able child showed the formation of equivalence classes, whereas each language-disabled child performed at chance level. Although correlational, the data provide support for the view that symbol use and the ability to form equiv-

alence classes are closely related. These results were found consistently with children across a range of chronological ages and mental-age scores. They included the youngest child (25 months) yet reported to show formation of an equivalence class, indicating that extensive language training or mastery is apparently not required for the formation of equivalence classes. In addition, for the first time, a clearly specified group of humans has been identified who apparently fail to form equivalence classes.

The results showed that this deficit cannot be explained on the basis of an inability to learn conditional discriminations per se. The retarded/no-language children required more trials and more prompts than the children in the other two groups to meet the mastery criterion in the conditional-discrimination training portion of the project. These data are consistent with reports on the acquisition of discriminations and conditional discrimina-
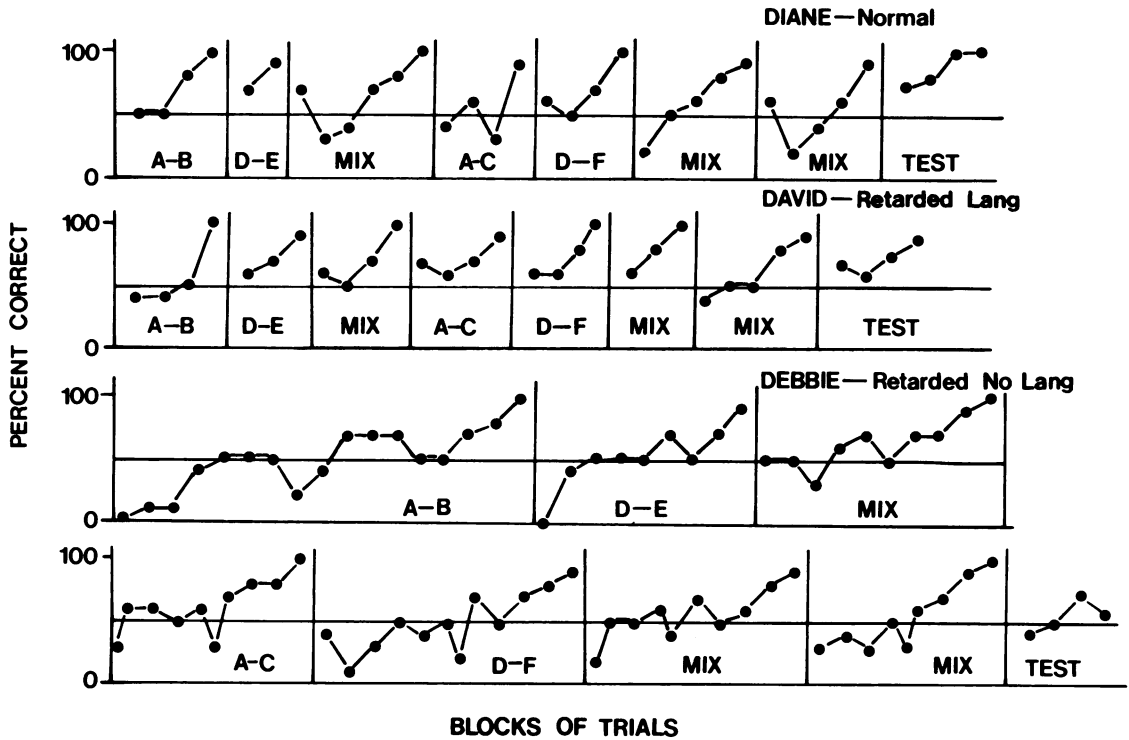
Fig. 5. Individual training and testing data for Diane, David, and Debbie (MA = approximately 19 months). The data are presented as the percentage of correct unprompted trials (vertical axis) across blocks of 10 trials (horizontal axis).

tions by retarded children with severely limited language abilities (e.g., Churchill, 1978; Lovaas, 1977; Routh, 1973). All the retarded/no-language children did acquire all the conditional discriminations, however, and were eventually able to respond consistently and correctly without prompting. None subsequently showed stimulus equivalence as a result of this successful training.

Supplementary data collected post hoc with retarded/no-language subjects and not reported here (see Devany, 1985) support the view that the failure of the retarded/no-language children on the equivalence test was not due to the use of unfamiliar, abstract, or nonsalient visual stimuli, nor to a failure to show generalized identity matching, but seemed to be related to the lack of symmetrical responding. Whenever symmetry has been shown in the equivalence literature, transitivity has also been found.

Different results might have been obtained had a more stringent mastery criterion been used than the one used here (9 of 10 responses

correct). Once the criterion was met, subjects proceeded immediately to the next phase of training or to testing. This may have precluded overtraining, which can alter the performance of severely handicapped individuals (Shover & Newsom, 1976). Had the training been greatly extended, the language-disabled children might have shown the development of equivalence classes. A major purpose, however, of overtraining would be to ensure that the baseline conditional discriminations were maintained. Because the subjects in the present study were tested immediately following mastery of the mixture of all four discriminations, maintained in the presence of a thinned schedule of reinforcement, it seems unlikely that the baseline discriminations would have deteriorated during the brief testing period.

It also seems unlikely that the responses of the children in the retarded/no-language group were extinguished during testing, as the level of noncontingent reinforcement in testing was matched to the level of contingent rein-
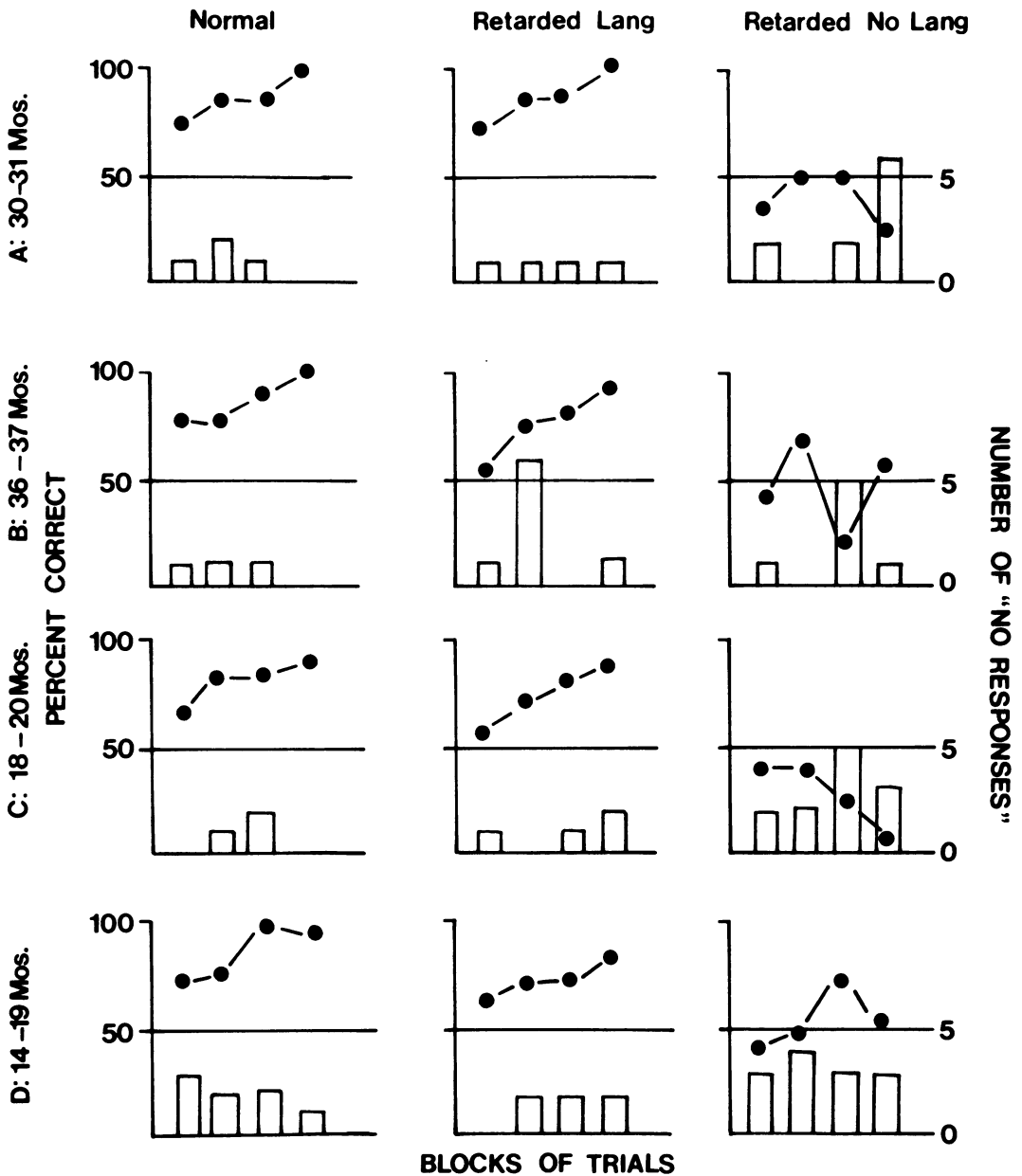
Fig. 6. Individual test data. Each graph presents the data for one child. Each column of graphs presents the data for those children in a particular language group. Each row of graphs presents the data for those children matched for mental age. The data (dots) are presented as the percentages correct (of all responses attempted) across blocks of 10 trials. The columns within each graph present the number of no responses that occurred within that block of 10 trials.

forcement used by the end of the training. Also, the testing phase might have been too short for these children to demonstrate equivalence classes, but this is not supported by trends in their testing data.

It is true that the retarded/no-language children differed from the other children in ways other than a failure to speak or sign. They required substantially more training to meet the mastery criterion and needed more prompting, for example. Matching by mental-age scores can only crudely equate children on

their level of intelligent behavior. These intelligence tests include some items that require verbal abilities and others that do not. Thus, a child without language skills could be matched on MA to a child with language skills only by having better developed nonverbal skills. The precise pattern of such abilities, and their causes, may be important in the formation of equivalence classes. Although language ability per se cannot be pinpointed as the source of the differences seen, these results make that possibility more plausible. It might be possible to disentangle this issue from that of retardation per se by examining equivalence-test performance of aphasic children with normal performance (nonverbal) IQ scores. Yet the important point here is that despite considerable heterogeneity within groups on a variety of measures, including trials to criterion, prompts, mental-age scores, and chronological age, there was no evidence of a relationship between these differences and equivalence-test performance within groups. Thus, there seems to be little basis for appealing to these variables as the source of the differences seen between groups. For example, the youngest normal child (Diane) took 273 trials to acquire the baseline conditional discriminations, and one of the higher MA retarded/no-language children (Barb) required 280 trials. However, Diane showed the acquisition of equivalence classes, but Barb did not.

It is not possible to say from these data whether the ability to form equivalence classes is a precursor of symbol use, a product of it, or if the two are both a reflection of the same process. It could be that the ability to form equivalence classes is a unique and distinct skill that itself is required for stimuli to be used symbolically. Conversely, language may be a distinct skill that in turn permits the formation of equivalence classes. Finally, it is possible that both the formation of equivalence classes and the acquisition of language are the result of other common processes.

Further analyses of the performances of very young developing children might help clarify this issue. If, for example, performance on an equivalence test is excellent before the child has acquired any labels, the argument that the ability to form equivalence classes is distinct (e.g., Sidman, 1986) and may itself lead to language acquisition would be strengthened.

Similarly, if successful language training also establishes equivalence-class formation in retarded children, the effect of language on equivalence classes would be implicated. If the two areas are essentially synonymous or if they both reflect common behavioral properties (such as the ability to respond in terms of arbitrary relations per se [e.g., Hayes, 1986; Hayes & Brownstein, 1985]), training in equivalence-class formation or its presumed underlying behavioral process should assist in language acquisition, and vice versa.

Improvement during testing. All of the language-able children showed improvement in test performance across blocks of test trials. This improvement in performance during testing has been seen in several studies (Lazar, Davis-Lang, & Sanchez, 1984; Sidman et al., 1985), although this study is the first in which the effect was seen in blocks of testing trials without conditional-discrimination training interspersed between test trials. The results of this project demonstrate that the interspersal of test and training trials is not necessary for the occurrence of improvement during testing and that improvement across the course of testing may be seen when testing trials are presented in a massed (as opposed to interspersed) format.

The reasons for this improvement are not clear. The improvement suggests that reinforcement of correct responding was occurring. In this project, however, responses were neither explicitly rewarded nor punished in the test phase. Any reinforcement that occurred in the test phase was unprogrammed. One possible source of reinforcement may have been subtle cues emitted by the experimenter (the Clever Hans effect; Sebeok, 1980). Although every effort was made to minimize any such cues, the possibility of subtle differential reactions cannot be disregarded. This improvement during testing, however, has been obtained repeatedly in studies in which automated equipment was used, which prevented the possibility of differential feedback (e.g., Sidman et al., 1985).

It has been suggested (Sidman et al., 1985) that this improvement occurs because the equivalence test itself provides a context in which the equivalence class is formed. In this view, the conditional-discrimination training provides the necessary history and the introduction of testing trials provides the necessary

context in which equivalence classes are formed. Alternatively, it may be that equivalence classes emerge during testing in part because humans have histories in which responding consistently had been more likely to be reinforced. In this view, subjects may discriminate the source of control over responses in each trial. At first, for example, a response may be at strength because one of the stimuli is pretty. Another may be at strength because one of the comparison stimuli looks like the sample stimulus. These bases, however, do not provide consistently strong responding in all trials and thus are discarded because reinforcement has rarely followed inconsistent responding in the past. Thus, perhaps a kind of response-contingent feedback is provided during testing—namely, the presence of responding based on a common source of control. Only responding controlled by an equivalence class will be at strength in every trial to which the subject is exposed, and thus over trials responses controlled by the equivalence class come to dominate other possible responses. This could be tested by including some irrelevant trials (with no correct answer based on equivalence) interspersed with the other testing trials. If this analysis is correct, this procedure should interfere with the formation of equivalence classes.

Receptive language and instructional control. The receptive language skills of the children in this project were not formally assessed. All of the children responded to simple instructions, such as "Sit down" and to consequences such as "No!," but the extent to which they could comprehend labels and other conversation is not known. Because this study used only novel visual stimuli, the extent to which receptive language skills might have contributed to success or failure in establishing equivalence classes is not clear. However, there are some data (Remington, 1985) indicating that receptive language may be an important variable in the formation of equivalence classes.

Because receptive language appears to be a matter of control by symbolic stimuli, stimulus equivalence may be relevant to rule-governed behavior. A growing body of literature (Baron & Galizio, 1983; Hayes, in press) suggests that human actions under the control of rules or instruction differ fundamentally from those under control of direct contingencies of

reinforcement. If equivalence classes have to do with language, we would also expect the control such stimuli exert over behavior to parallel the control that instructions seem to exert over behavior in humans. The available evidence, while still limited, indicates that this is what occurs. For example, if the spoken and written words "men" and "boys" and pictures of males are part of an equivalence class ("males"), and a boy learns to enter the "men's room" at a restaurant, he may also come to enter the "boy's room" at another, or the room with a drawing of a stick person with pants at a third, without explicit training to do so. As a parallel, after an equivalence class (ABC) is formed, if one member (A) becomes discriminative for a response, then B and C will also become discriminative for the same response (Hayes, Brownstein, Devany, Kohlenberg, & Shelby, 1985; see also Lazar, 1977). To take another example, if someone is told that the word for "good" in French is "bon," and the word for "bon" in Spanish is "bueno," and if "good" functions as a conditioned reinforcer for behavior, it seems likely that "bueno" will now also do so, without "bueno" ever having been paired with positive reinforcers. Similar results have been found with equivalence classes (Hayes et al., 1985).

Thus, it may be possible that instructional control is based fundamentally on stimulus equivalence as a behavioral process (Hayes, 1986; Hayes & Brownstein, 1985). For example, consider a dog's approach to its owner when commanded "come here." The action of the dog seems similar to the behavior of a human being when given the same command. There appears to be nothing special about the control exerted by verbal stimuli over a human as compared with the control exerted over an infrahuman by discriminative stimuli more generally. It is possible, however, that the source of control over the action of the dog and human is different. The words "come here" may be effective for the dog because in the past the probability of reinforcement for approach was higher in their presence than in their absence. The human, however, may be responding not because of a direct history of this sort, but in part because these stimuli participate in equivalence classes established by the verbal community. Their discriminative effect may be only indirectly tied to a direct history of reinforcement involving other

members of the class, or to additional classes related to the class. If so, rules and instructions may exert their control through very different processes than do typical discriminative stimuli.

The data from this project support the view that there is a relation between the ability to form equivalence classes and language. Specifically, the language-disabled children were shown to be unable to form equivalence classes under conditions in which very young normal and language-able retarded children were easily able to do so even though they mastered the requisite conditional discriminations. This connection between language and stimulus equivalence may lead to the development of training techniques for the remediation of language and generalization deficits in developmentally disabled populations.

In the thirty years since the publication of *Verbal Behavior* (Skinner, 1957), empirical progress in the behavior-analytic understanding of language has been disappointing. This may be, in part, because research has focused on the behavior of the speaker, creating notable difficulties for a behavior-analytic approach in the areas of response definition and measurement (Hayes & Brownstein, 1984). If the present analysis is correct, the study of stimulus equivalence provides another, possibly more fruitful, avenue for the study of language phenomena.

REFERENCES

Baron, A., & Galizio, M. (1983). Instructional control of human operant behavior. *Psychological Record, 33,* 495–520.

Churchill, D. W. (1978). *Language of autistic children.* Washington, DC: V. H. Winston.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics.* New York: Harcourt, Brace, Jovanovich.

D'Amato, M. R., Salmon, D. P., Loukas, E., & Tomie, A. (1985). Symmetry and transitivity of conditional relations in monkeys (*Cebus apella*) and pigeons (*Columba livia*). *Journal of the Experimental Analysis of Behavior, 44,* 35–47.

Devany, J. M. (1985). *Stimulus equivalence and language development in children.* Unpublished doctoral dissertation, University of North Carolina–Greensboro.

Dixon, M. (1976). Teaching conceptual classes with receptive label training. *Acta Symbolica, 7,* 17–35.

Dixon, M., & Spradlin, J. (1976). Establishing stimulus equivalences among retarded adolescents. *Journal of Experimental Child Psychology, 21,* 144–164.

Edwards, C. A., Jagielo, J. A., & Zentall, T. R. (1983). "Same/different" symbol use by pigeons. *Animal Learning & Behavior, 11,* 349–355.

Ferguson, G. A. (1976). *Statistical analysis in psychology and education* (4th ed.). New York: McGraw-Hill.

Fouts, R. S., Chown, B., & Goodin, L. (1976). Transfer of signed responses in American Sign Language from vocal English stimuli to physical object stimuli by a chimpanzee (*Pan*). *Learning and Motivation, 7,* 458–475.

Gast, D. L., VanBiervliet, A., & Spradlin, J. E. (1979). Teaching number–word equivalences: A study of transfer. *American Journal of Mental Deficiency, 83,* 524–527.

Hayes, S. C. (1986). The case of the silent dog—Verbal reports and the analysis of rules. A review of Ericcson & Simon's, *Protocol Analysis: Verbal Reports as Data. Journal of the Experimental Analysis of Behavior, 45,* 351–363.

Hayes, S. C. (Ed.). (in press). *Rule-governed behavior: Cognition, contingencies, and instructional control.* New York: Plenum Press.

Hayes, S. C., & Brownstein, A. J. (1984). Verbal behavior: Is the human operant laboratory an ideal place to begin? *Experimental Analysis of Human Behavior Bulletin, 2,* 11–13.

Hayes, S. C., & Brownstein, A. J. (1985, May). *Verbal behavior, equivalence classes and rules: New definitions, data, and directions.* Paper presented at the annual meeting of the Association for Behavior Analysis, Columbus, OH.

Hayes, S. C., Brownstein, A., Devany, J., Kohlenberg, B., & Shelby, J. (1985, May). *Stimulus equivalence and the acquisition of meaning.* Paper presented at the annual meeting of the Association for Behavior Analysis, Columbus, OH.

Herman, L. M., & Thompson, R. K. R. (1982). Symbolic, identity, and probe delayed matching of sounds by the bottlenosed dolphin. *Animal Learning & Behavior, 10,* 22–34.

Kendall, S. B. (1983). Tests for mediated transfer in pigeons. *Psychological Record, 33,* 245–256.

Lashley, K. S. (1938). Conditional reactions in the rat. *Journal of Psychology, 6,* 311–324.

Lazar, R. (1977). Extending sequence-class membership with matching to sample. *Journal of the Experimental Analysis of Behavior, 27,* 381–392.

Lazar, R. M., Davis-Lang, D., & Sanchez, L. (1984). The formation of visual stimulus equivalences in children. *Journal of the Experimental Analysis of Behavior, 41,* 251–266.

Lovaas, O. I. (1977). *The autistic child.* New York: Irvington.

Mackay, H. A., & Sidman, M. (1984). Teaching new behavior via equivalence relations. In P. H. Brooks, R. Sperber, & C. McCauley (Eds.), *Learning and cognition in the mentally retarded* (pp. 493–513). Hillsdale, NJ: Erlbaum.

Nelson, K. (1974). Concept, word, and sentence: Interrelations in acquisition and development. *Psychological Review, 81,* 267–285.

Nissen, H. W. (1951). Analysis of a complex conditional reaction in chimpanzee. *Journal of Comparative and Physiological Psychology, 44,* 9–16.

Premack, D. (1976). *Intelligence in ape and man.* Hillsdale, NJ: Erlbaum.

Remington, R. (1985, April). *Is signing a substitute for speech with retarded children?* Paper presented at the University of North Carolina–Greensboro.

Routh, D. K. (Ed.). (1973). *The experimental psychology of mental retardation.* Chicago: Aldine.

Savage-Rumbaugh, E. S. (1984). Verbal behavior at a procedural level in the chimpanzee. *Journal of the Experimental Analysis of Behavior, 41,* 223–250.

Sebeok, T. (1980). Looking in the destination for what should have been sought in the source. In T. A. Sebeok & J. Umiker-Sebeok (Eds.), *Speaking of apes* (pp. 407–428). New York: Plenum Press.

Shover, L. R., & Newsom, C. D. (1976). Overselectivity, developmental level, and overtraining in autistic and normal children. *Journal of Abnormal Child Psychology, 4,* 289–298.

Sidman, M. (1971). Reading and auditory–visual equivalences. *Journal of Speech and Hearing Research, 14,* 5–13.

Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 213–245). Hillsdale, NJ: Erlbaum.

Sidman, M., Cresson, O., Jr., & Willson-Morris, M. (1974). Acquisition of matching to sample via mediated transfer. *Journal of the Experimental Analysis of Behavior, 22,* 261–273.

Sidman, M., Kirk, B., & Willson-Morris, M. (1985). Six-member stimulus classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behavior, 43,* 21–42.

Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., & Carrigan, P. (1982). A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children. *Journal of the Experimental Analysis of Behavior, 37,* 23–44.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37,* 5–22.

Skinner, B. F. (1957). *Verbal behavior.* New York: Appleton-Century-Crofts.

Spradlin, J. E., Cotter, V. W., & Baxley, N. (1973). Establishing a conditional discrimination without direct training: A study of transfer with retarded adolescents. *American Journal of Mental Deficiency, 77,* 556–566.

Spradlin, J. E., & Dixon, M. H. (1976). Establishing conditional discriminations without direct training: Stimulus classes and labels. *American Journal of Mental Deficiency, 80,* 555–561.

VanBiervliet, A. (1977). Establishing words and objects as functionally equivalent through manual sign training. *American Journal of Mental Deficiency, 82,* 178–186.

Wetherby, B., Karlan, G. R., & Spradlin, J. E. (1983). The development of derived stimulus relations through training in arbitrary-matching sequences. *Journal of the Experimental Analysis of Behavior, 40,* 69–78.